

## ABSTRACT

---



## CONTENTS

---

## LIST OF FIGURES

---

## LIST OF TABLES

---

## LISTINGS

---

## ACRONYMS

---

NIND Natural Image Noise Dataset

BM<sub>3</sub>D Block-matching and 3D filtering [\[4\]](#)

## INTRODUCTION

---

Photographic image noise occurs as a camera sensor's ISO sensitivity increases to capture an image faster than it would in ideal conditions ("base ISO" sensitivity).<sup>1</sup> A fast shutter speed is often necessary even though there is insufficient light, for instance with handheld photography where a slow shutter speed results in blur caused by the camera shake, or when a dynamic subject results in motion blur. Increasing the ISO setting is akin to linearly amplifying the value measured on each sensor cell. A small initial value that is amplified is less accurate and more prone to errors; this amplified value in turn makes up photographic noise.

Denoising is typically seen as the inverse problem of recovering the latent clean image from its noisy observation [5].

Photographic development usually incorporates a conventional denoising method such as non-local means [nlm] or wavelets-based methods [wavelets-denoising], or a combination thereof, sometimes making use of prior information about a specific sensor's response to different ISO noise values. [darktable-denoising] Block-matching and 3D filtering [4] (BM3D) is a more powerful denoising method which has been used to benchmark newer techniques [all-bm3d-benchmarks], though it is seldom used in off-the-shelf software (presumably because of its high computational cost and the need to specify a  $\gamma$  noise value parameter).

The use of deep learning to solve the denoising problem by directly generating the latent clean image, or in some cases recreating the noise and subtracting it from the observed image [9], has been an active research area. However, while the synthetic results show state-of-the-art performance, testing on real data indicates that neural network-based solutions do not exceed the performance offered by BM3D [6]. It appears that neural networks simply learn the applied noise distribution and that ISO noise may involve additional transformations such as color distortions and loss of detail.

Some specialized work has shown that neural networks obtain state-of-the-art performance when trained with real data [3][10]. We sought to assess the potential of deep learning applied to the denoising problem by expanding on this previous work through a dataset of images produced with various levels of ISO noise. This dataset can be used to train neural network models for general purpose denoising of high quality images.

---

<sup>1</sup> We often make references to ISO noise because increased ISO sensitivity is the main cause of noise, but it should be noted that there are other factors affecting the magnitude of noise acquired by the image sensor.

Our work introduces an open dataset of DSLR-like<sup>2</sup> image sets with various levels of real noise caused by the digital sensor’s increased ISO sensitivity. The dataset is large enough to be used for training and varies in content in order to model a great variety of scenes. Each scene was captured at multiple noise levels, with an average of 6 images per set, such that a model may be trained for blind denoising on the base ISO as well as beyond the highest ISO value of the camera by feeding it crops that have a random noise value.

The images in a set are all pixel-aligned. Some of the scenes include multiple ground-truth images which may be sampled at random during training; these would prevent the model from learning to reconstruct the random noise it has seen on one ground-truth, thus making it more difficult to overfit the noise. Overexposed areas are avoided in the ground-truth images because details are lost when the sensor is saturated and this could potentially give an advantage to higher ISO images in which the sensor is not necessarily saturated (notably in ISO-invariant cameras and high ISO pictures that are brightened using software).

The dataset is published in sRGB format on Wikimedia Commons, which is an open-platform that promotes continuous discussion and contribution.

We trained several convolutional neural network architectures presented in the literature, namely DnCNN [9], RedNet [5], and U-Net [7], and analysed different proposed methods such as DnCNN [9]’s residual learning.

In most tests a U-Net model was trained with the Natural Image Noise Dataset (NIND) and validated over increasing ISO values; test sets were taken with the same Fujifilm X-T1, as well as a Canon EOS 500D DSLR camera to assess the generalization. We also attempted to combine other datasets with our training data to assess the potential loss in performance caused by generalization. Finally we investigated the use of Generative Adversarial Networks[p2p] to solve issues such as unnatural face smoothing.

---

<sup>2</sup> We define a DSLR-like camera as one produced with an APS-C (25.1×16.7 mm) or larger sensor such as those present in most DSLR and mirrorless cameras

## DATASET

---

The Dataset chapter goes over some existing image noise datasets before delving into the Natural Image Noise Dataset.

### 2.1 RELATED WORK

The following works feature datasets made of ground-truth / noisy image sets. The static scenes approach is necessary to directly compare the level of degradation using a loss function such as the structural similarity index (SSIM) [8] or the mean square error (MSE).

#### 2.1.1 Darmstadt Noise Dataset

The Darmstadt Noise Dataset (DND) [6], containing 50 pairs of noisy-clean images from four cameras, was developed for the purpose of validating denoising algorithms using real data. Synthetic noise is typically used to train and test models, but it had been unclear whether the reported synthetic results translated to real improvements. Plötz and Roth showed that many modern denoising methods do not perform well on real data and that BM3D [bm3d], which was published in 2006, remains one of the best performing methods [6]. RENOIR [2] is a similar dataset that was published prior to the DND; however, Plötz and Roth noted spatial misalignments that reduced its effectivity. We have additionally found that the light sometimes differs between images in the same scene and that some photographs exhibit significant raw overexposure.

#### 2.1.2 Learning to See in the Dark

See-in-the-Dark (SID) [3] is an image noise dataset that is large enough for training and, to our knowledge, was used in the first successful attempt at denoising images using real image noise. This dataset focused on very low-light photography where the camera-generated JPEG appears black. The authors used a U-Net network architecture to create an end-to-end RAW-to-JPEG pipeline that produces realistic colors, improving on standard processing and BM3D denoised images which still suffer from color bias at high ISO. Our work differs from SID in that we aimed to train a general purpose ("blind") denoiser rather than one that handles a specific condition, such as extremely low light images. We chose to work in sRGB space because handling the whole RAW-to-sRGB pipeline removes some information which

may otherwise be useful to the author during development. Moreover, one dataset can then be used with different types of color filter arrays.

### 2.1.3 Smartphone Image Denoising Dataset

The Smartphone Image Denoising Dataset (SIDDD) [1] is comprised of 10 scenes \* 5 cameras \* 4 conditions \* 150 images, totalling 30000 images. This dataset aims to address the problem of smartphone image denoising, where the small sensor and aperture size causes noticeable noise even in pictures taken at base ISO. Further processing is thus applied to create ground-truth images out of many images. This method of creating ground-truth images is not entirely relevant for denoising images captured with larger sensors because a single image taken at base ISO on a DSLR-like camera is clean enough to work as ground-truth for training purposes.

## 2.2 NATURAL IMAGE NOISE DATASET

*DRAFTy-as-heck  
sentence*

The Natural Image Noise Dataset fills the gap for an image noise dataset of ground-truth - noisy sets that targets cameras equipped with larger sensors<sup>2</sup> and which is large and varied enough (in both noise values and content types) to train a denoising model for ISO-blind denoising.

*FIXME long run-on  
sentence*

Here we outline the physical setup required to capture image sets for the NIND, summarize its content, explain the software processing and validation requirements, and describe its publication aspects such that others that wish to do so may also contribute.

### 2.2.1 Capture

We captured several images per static scene; at least one ground-truth taken with the camera's lowest ISO setting and several images taken with increasing ISO settings and consequent decreasing shutter speed in order to match the original exposure value. Scenes were captured using a camera affixed to a tripod and controlled with a wireless remote control to avoid shifting the setup position. We ensured that the ground was stable, wind would not cause any change in the scene, lighting did not vary between shots, and no area was overexposed in the ground-truth images. Overexposure occurs when the sensor is saturated; on a ground-truth image this would potentially benefit the high ISO images because less light is captured with a faster shutter speed, therefore, the higher ISO images may not be overexposed and thus may contain detail that is not present in the overexposed ground-truth. The aperture remained the same on all shots and the focus was set manually so that it would not automatically adjusted for each frame.



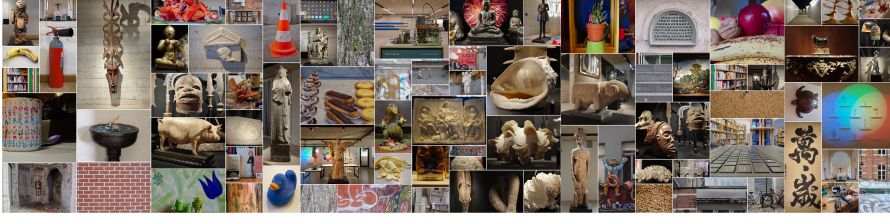


Figure 2.1: Sample ground-truth images from the Natural Image Noise Dataset.

ISO value	100	200	250	320	400	500	640	800	1000	1250	1600	2000	2500	3200	4000	5000	6400	High	Scenes	Images
Fujifilm X-T1		105	11	8	18	13	16	25	16	9	12	9	19	25	17	19	91	123	90	536
Canon C500D	14	10			10			10			9			11				10	11	74
Total	15	116	11	8	29	13	16	36	16	9	22	9	19	37	17	19	91	133	101	616

Table 2.1: Dataset content

A base ISO image (ISO200 on the Fujifilm X-T1) was always taken at least once, along with the camera’s highest ISO setting (ISO6400 on the X-T1). Several images were taken with different intermediate ISO values such that the ISO settings varied across each scene. We often also took images that we categorized as “High ISO,” which consisted of the highest ISO value and increased shutter speed. “High ISO” images result in dark frames which are then correctly exposed using software. We often tried to match shutter speeds that would be useful to denoise, such as  $1/60s$  for handheld photography,  $1/15s$  for devices equipped with optical image stabilization, and  $1/1000s$  or faster for high-speed photography. We ensured that every ISO value was well represented in order to train models effective at blind denoising. On average, six images were produced per scene and some scenes featured multiple ground truth images which could be used in training to help prevent over-fitting.

### 2.2.2 Content

Many objects were captured in museums where subjects are plentiful (albeit we had to be mindful of copyright restricted material) and for which a denoising application would be highly relevant because indoor handheld pictures require a high ISO sensitivity. However, initial tests found that using images taken only indoors did not provide the variety needed to create a model that generalized well across all conditions, as natural colors were sometimes off in outdoor and brightly lit applications. We thus captured natural objects with vibrant colors (such as food items and plant-life) as well as outdoor scenes where the shutter speed could be taken as fast as  $1/13000s$  using a digital shutter. We made an effort to include some text because it is prevalent, yet we expect a model would not be able to guess how to reconstruct it (Figure ?? shows the resulting denoised text), and we tried to make the images pleasant to look at in order to enhance the

time users would spend looking at them. Most of the NIND images were captured on a Fujifilm X-T1 mirrorless camera, which uses a 23.6 x 15.6 mm X-Trans sensor. A Canon EOS 500D DSLR camera, featuring a 22.3 x 14.9 mm standard Bayer sensor, was used to capture images that could be used to validate the generalization. The content of the dataset is summarized in Table ?? and a subset of the X-T1 pictures are shown in Figure ??.

### 2.2.3 *Software processing*

We processed the dataset images using darktable [**darktable**] (an open source development software) for raw-to-sRGB development. Our development steps are similar to those we would apply to a standard picture but no sharpening was applied, as this greatly amplifies noise and is typically applied last in the pixel pipeline (we can expect users to apply sharpening to the generated clean image without any perceptible loss). We used darktable’s automatic exposure mode to match a fixed percentile target on the histogram and calculate the required exposure compensation for all images in a set. Likewise, we ensured that the white balance was identical and all development steps were copied over to the entire scene. (The exposure percentile and white balance settings are fixed within a scene but vary across different scenes.) The raw overexposed indicator was used to verify that no overexposed areas were present in the base ISO image and if any were detected then it was cropped out or the scene was discarded. The images were visually inspected to detect slight variations in light, the introduction of foreign objects such as insects, or any movement, which also resulted in cropping or discarding images. The ground-truth images must be at least as sharp as their noisy counterparts; this is sometimes not the case due to slight movements on longer exposures. The remaining images were saved in either high-quality (98 to 100) 8-bit JPEG or lossless 16-bit PNG.

The last step of development is to use Hugin’s align\_image\_stack tool [**hugin**] to ensure that all images in a set are perfectly pixel-aligned. The tool will usually return the same image size as the input, in which case the whole image set can safely be used. When a difference is detected then the tool will automatically align the set and we visually analyzed whether the result was acceptable or the movement caused a change in perspective, in which case the outlier images were discarded. Some noisy images cannot easily be matched to the scene; possible solutions are to denoise these images in order to check the alignment or to take a cleaner image afterward and assume that the middle images are consistent with the previous and next ones.

#### 2.2.4 Publishing

The Natural Image Noise Dataset is published on Wikimedia Commons ([https://commons.wikimedia.org/wiki/Natural\\_Image\\_Noise\\_Dataset](https://commons.wikimedia.org/wiki/Natural_Image_Noise_Dataset)), an online repository of free-use images and other digital media. Wikimedia Commons hosts media content for all Wikimedia projects and its scope is limited only by the content having some educational value (broadly meaning “providing knowledge; instructional or informative”). As such, we believe it is a fitting platform for the publication of a research dataset.

One key advantage of using Wikimedia Commons is its collaborative aspect. Anyone is allowed to add images to the dataset, modify existing images (for example to fix a spatial misalignment), and discuss the content (through the discussion page provided for each file, category, and the dataset itself). The collaborative aspect also includes a “Quality images candidates” page [qic] where users assess the technical quality of a submitted image and may promote it to a “Quality image” standing. Many of the ground-truth images have gone through this process and were promoted through human assessment. The same process was also used to validate the trained model, which ended up in a positive assessment since even the denoised dynamic ISO6400 picture presented in Figure ?? was among the promoted images.

On the technical side, Wikimedia Commons preserves images as they are uploaded; JPEG images are not recompressed, 16-bit lossless TIFF and PNG images are allowed, and the metadata is kept. Thumbnails are generated and images may be visualized before being downloaded in full resolution, and the download may include a select subset instead of the whole dataset. A customizable download script is provided on the dataset’s page for convenient retrieval. Even though files can be overwritten, every file uploaded on Wikimedia Commons is kept forever therefore specific snapshots of the dataset can be made by including the files’ revision in the download script and getting a specific version ID (or commit hash) of the download script.

It is also important to note the usefulness of the data outside of a denoising dataset context. Many images, such as those depicting artifacts displayed in churches and museums, have encyclopedic value and the ground-truth images present in our dataset are of higher quality than most previously available images depicting such artifacts. By publishing the dataset on Wikimedia Commons, these ground-truth images may be used in Wikipedia articles directly<sup>1</sup>. This may be a motivating factor to those wishing to contribute.

---

<sup>1</sup> E.g. [Bobo people](#), [Bombardment of Brussels](#), [Dengese people](#)



## NEURAL NETWORK ARCHITECTURES

---

### 3.1 CONVOLUTIONAL NEURAL NETWORKS

The denoising methods we use are based on convolutional neural networks.

shared weights

*DRAFTy-as-heck*

### 3.2 NETWORK ARCHITECTURES

- present dncnn, rednet, unet,



## GENERATIVE ADVERSARIAL NETWORKS

---

Introduce

### 4.1 TYPES OF GANS

CycleGAN, does not apply here because

### 4.2 CHOSEN GAN

cGAN and GAN PatchGAN The generator is the same UNet architecture used in previous





RESULTS

---



## CONCLUSION

---

### 6.1 FUTURE WORK

- Discriminator loss
  - multi-scale (pix2pixHD)
  - train a general network with specialized dataset



## BIBLIOGRAPHY

---

- [1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S. Brown. “A High-Quality Denoising Dataset for Smartphone Cameras.” In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018.
- [2] Josue Anaya and Adrian Barbu. “RENOIR – A dataset for real low-light image noise reduction.” In: *Journal of Visual Communication and Image Representation* 51 (Sept. 2014). DOI: [10.1016/j.jvcir.2018.01.012](https://doi.org/10.1016/j.jvcir.2018.01.012).
- [3] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. “Learning to See in the Dark.” In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018.
- [4] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. “Image denoising with block-matching and 3D filtering.” In: *Proceedings of SPIE - The International Society for Optical Engineering* 6064 (Feb. 2006), pp. 354–365. DOI: [10.1117/12.643267](https://doi.org/10.1117/12.643267).
- [5] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. “Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections.” In: *Proc. Advances in Neural Inf. Process. Syst.* 2016.
- [6] Tobias Plotz and Stefan Roth. “Benchmarking Denoising Algorithms With Real Photographs.” In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [7] O. Ronneberger, P. Fischer, and T. Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Vol. 9351. LNCS. (available on arXiv:1505.04597 [cs.CV]). Springer, 2015, pp. 234–241. URL: <http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a>.
- [8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. “Image Quality Assessment: From Error Visibility to Structural Similarity.” In: *IEEE Transactions on Image Processing* 13 (Apr. 2004), pp. 600–612. DOI: [10.1109/TIP.2003.819861](https://doi.org/10.1109/TIP.2003.819861).
- [9] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. “Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising.” In: *IEEE Transactions on Image Processing* 26.7 (2017), pp. 3142–3155.

- [10] Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, and Scott Howard. "A Poisson-Gaussian Denoising Dataset with Real Fluorescence Microscopy Images." In: *CoRR* abs/1812.10366 (2018). arXiv: 1812.10366. URL: <http://arxiv.org/abs/1812.10366>.